DOCUMENT RESUME

ED 480 045 CG 032 618

AUTHOR	Goldsmith, Sharon M.
TITLE	Lost in Translation: Issues in Translating Tests for Non-English Speaking, Limited English Proficient, and Bilingual Students.
PUB DATE	2003-08-00
NOTE	22p.; In: Measuring Up: Assessment Issues for Teachers, Counselors, and Administrators; see CG 032 608.
PUB TYPE	Information Analyses (070)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Bilingualism; Educational Assessment; *Educational Testing; *English (Second Language); High Stakes Tests; Limited English Speaking; Test Construction; Test Use; *Testing Problems; *Translation

ABSTRACT

The need to conduct assessments in languages other than English is growing rapidly. In addition to the rising number of children who do not speak English as their language at home, the number of different languages spoken by children in public schools is also increasing rapidly. Shifting demographics strongly support the need to increase the number of assessments that are available in languages other than English. This chapter highlights several different reasons to provide test translations, including the increased emphasis on assessment, particularly large-scale, high-stakes assessments, in public schools. Also discussed are problems in test translation and alternatives to test translation. (Contains 20 references.) (GCP)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 480 045

Lost in Translation: Issues in Translating Tests for Non-English Speaking, Limited English Proficient, and Bilingual Students

By
Sharon M. Goldsmith

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

CG032618

Chapter 10
Lost in Translation
Issues in Translating Tests for Non-English-Speaking, Limited English Proficient, and Bilingual Students

Sharon M. Goldsmith

The need to conduct assessments in languages other than English is growing rapidly. According to Geisinger and Carlson (1992), 15 to 20 percent of school-age children speak a foreign language at home and do not speak English as their primary language. In addition to the rising number of children who do not speak English as their primary language at home, the number of different languages spoken by children in public schools is also increasing rapidly. Bracken and McCallum (1999) conducted a meta-analysis of the studies examining the languages used in U.S. public schools. They reported that children enrolled in the Chicago public schools alone speak one or more of 200 languages; 1.4 million children in the California public schools speak one or more of 150 languages. Several school districts, including Scottsdale, Arizona, Palm Beach, Florida, and Prince William County, Maryland, report between 40 and 80 different languages spoken by children attending their schools. Even in small communities, students speak a large number of languages. Bracken and McCallum (1999) reference a study that reported 30 languages being spoken in a single small high school in the Washington state rural community of Tukwila.

The Case for Test Translation

Shifting demographics strongly support the need to increase the number of assessments that are available in languages other than English. There are several additional reasons to provide test translations, including the increased emphasis on assessment, particularly large-scale, high-stakes assessments, in public schools.

The Council of Chief State School Officers' (2001) survey on public school assessments reports that in 1999, 48 states required statewide assessments in math and language arts, 33 required statewide

assessments in science, and 29 required assessments in social studies. These assessments can take a variety of forms. Although the majority of statewide assessments rely on multiple-choice responses, other formats, such as extended response, short-answer, portfolio, and performance, are also common.

These assessments are high stakes for the student in that decisions regarding promotion to a higher grade or graduation from high school may be dependent on the student's performance on these tests. These assessments are also high stakes for the teachers and school administrators because student performance may affect decisions regarding tenure, compensation, and eligibility for state and federal funding. There is pressure on both students and schools to perform well on these assessments.

The majority of assessment procedures are highly language dependent. Demonstrating knowledge of almost any subject matter is dependent on the ability to read and answer questions in English. Even alternative assessment procedures require the ability to follow directions provided in English. The impact of language skill on success in schools cannot be overemphasized. The low performance on high-stakes assessments of students with limited English proficiency (LEP) is well documented.

The *Standards for Educational and Psychological Testing* (1999), issued by the American Educational Research Association, the American Psychological Association, and the National Council for Measurement in Education, provides guidance regarding the construction, evaluation, and use of tests. Standard 11.22 states:

> When circumstances require that a test be administered in the same language to all examinees in a linguistically diverse population, the test user should investigate the validity of the score interpretations for test takers believed to have limited proficiency in the language of the test. The achievement, abilities and traits of examinees who do not speak the language of the test as their primary language may be seriously mismeasured by the test. (p. 118)

Historically, many school districts have addressed the low performance of students with limited English by simply exempting them from participating in these assessment programs. These exemptions were designed to maintain high average test results for districts by not

having the mean scores influenced by the scores of those students who statistically do not do as well, particularly students with special needs and those who are in the linguistic minority. Federal legislation now prohibits schools from simply exempting students and requires schools to provide appropriate test accommodations instead. The Individuals With Disabilities Act (IDEA) requires that assessments be administered in the student's native language or in the language used in the student's home.

The need to provide appropriate accommodations to linguistic minority students is also being driven by new legislation such as the No Child Left Behind Act (NCLB) signed by President Bush as part of the reauthorization of the Elementary and Secondary Education Act (ESEA). ESEA will increase the accountability of states for the academic performance of public school students. Among other requirements, states will be required to establish performance standards against which students will be measured. The NCLB also requires states to include more students in assessment programs by creating appropriate accommodations for them. The act will consolidate funding for bilingual education and will require states to test those students with limited English proficiency who have had at least three years of schooling in the United States.

Several professional societies that are concerned with assessment issues have taken the position that assessments are to be conducted in the student's primary language. For example, the American Speech-Language-Hearing Association (ASHA, 1985), in a technical report on the clinical management of communicatively handicapped minority populations, states that assessment should be conducted in the client's primary language.

Therefore, the question for many states and school districts is not whether to provide accommodations for students in the linguistic minority but how. Several researchers, including Figueroa (1990) and others, have suggested that the best accommodation is to assess linguistic minorities in their native language. Several states do offer translations of tests in several languages. For example, New York state offers its high school graduation test, the Regents Competency Examination, in 20 languages. Even small states offer different language versions of statewide tests. Rhode Island, for example, offers its state test for grades 4, 8, and 10 in four languages: Spanish, Laotian, Portuguese, and Cambodian.

Guidelines for Test Translation

Translating tests is a complicated process. Several guidelines should be followed to achieve a quality translation, such as those put forth by the International Test Commission's *International Guidelines for Test Use* (1999):

> When testing in more than one language (within or across countries), competent test users will make all reasonable efforts to ensure that:
>
> - Each language or dialect version has been developed using a rigorous methodology meeting the requirements of best practice;
> - The developers have been sensitive to issues of content, culture and language;
> - Test administrators can communicate clearly in the language in which the test is to be administered;
> - The test taker's level of proficiency in the language in which the test will be administered is determined systematically and the appropriate language version is administered or bilingual assessment is performed, if appropriate. (p. 13)

Standard 9.7 of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) provides additional guidance:

> When a test is translated from one language to another, the methods used in establishing the adequacy of the translation should be described, and empirical and logical evidence should be provided for score reliability and the validity of the translated test's score inference for the uses intended in the linguistic groups to be tested.
>
> For example, if a test is translated into Spanish for use with Mexican, Puerto Rican, Cuban, Central American, and Spanish populations, score reliability and the validity of test score inferences should be established with members of each of these groups separately where feasible. In addition, the test translation methods used need to be described in detail. (p. 99)

Standard 9.6 states:

> When a test is recommended for use with linguistically diverse test takers, test developers and publishers should provide the information necessary for appropriate test use and interpretation.
> Test developers should include in test manuals and in instruction for score interpretation explicit statements about the applicability of the test with individuals who are not native speakers of the original language of the test. However, it should be recognized that test developers and publishers seldom will find it feasible to conduct studies specific to the large number of linguistic groups found in certain countries. (p. 99)

Standard 9.4 states:

> Linguistic modifications recommended by test publishers, as well as the rationale for the modifications, should be described in detail in the test manual.
> Linguistic modifications may be recommended for the original test in the primary language or for an adapted version in a secondary language, or both. In any case, the test manual should provide appropriate information regarding the recommended modifications, their rationales, and the appropriate use of scores obtained using these linguistic modifications. (p. 98)

Test translation requires much more than translating the words on a test from one language to another. It requires constructing an entirely new test. It requires making sure that the semantic content of the test and the concepts used are culturally appropriate and likely to be understood by the test taker. For example, in the widely used Peabody Picture Vocabulary Test that is used to assess language understanding (i.e., receptive vocabulary skills), the test taker is shown a page with four pictures and asked to point to the correct picture as it is named. Several of the pictures are of items or scenes that are familiar in U.S. middle-class culture but would not be familiar in other cultures or environments. Simply translating the verbal stimulus (the word to be identified) into the child's language is not sufficient to measure language understanding accurately. The pictures themselves as well as the

vocabulary being tested would need to be appropriate for both the cultural and the linguistic environment.

Additionally, the translated test, even with appropriate linguistic and cultural modifications to the content, would need to be subjected to new analyses of reliability, validity, and scoring norms against the population for which the test has been translated. Standard 9.1 of the *Standards for Educational and Psychological Testing* (AERA et al., 1999) emphasizes the need to establish reliability and validity for a translated test:

> Testing practice should be designed to reduce threats to the reliability and validity of test score inferences that may arise from language differences. (p. 97)

Standard 9.2 states:

> When credible research evidence reports that test scores differ in meaning across subgroups of linguistically diverse test takers, then to the extent feasible, test developers should collect for each linguistic subgroup studied the same form of validity evidence collected for the examinee population as a whole. (p. 97)
>
> Linguistic subgroups may be found to differ with respect to what test content is appropriate, how their test responses are internally structured, how their test scores relate to other variables, and what response processes individual examinees employ. Any such findings need to receive due consideration in the interpretation and use of scores as well as in test revisions. There may also be legal or regulatory requirements to collect subgroup validity evidence. Not all forms of evidence can be examined separately for members of all linguistic groups. The validity argument may rely on existing research literature, for example, and such literature may not be available for some populations. For some kinds of evidence, separate linguistic subgroup analyses may not be feasible due to the limited number of cases available. Data may sometimes be accumulated so that these analyses can be performed after the test has been in use for a period of time. It is important to note that this standard calls for more than representativeness in the selection of samples used for validation or norming studies.

> Rather, it calls for separate, parallel analyses of data for members of different linguistic groups, sample sizes permitting. If a test is being used while such data are being collected, then cautionary statements are in order regarding limitations on the interpretations based on test scores.

Standard 9.9 discusses establishing and interpreting test scores in translated tests:

> When multiple language versions of a test are intended to be comparable, test developers should report evidence of test comparability.
> Evidence of test comparability may include, but is not limited to, evidence that the different language versions measure equivalent or similar constructs, and that score reliability and the validity of inferences from scores from the two versions are comparable. (p. 99)

Although the guidelines previously outlined address the philosophical principles of what is required in quality test translation, other guidelines focus on specific procedures that should be followed in test translation. Many of these guidelines are issued by test developers who expect their tests to be translated into other languages, often for use in other countries. The developers are interested in making sure that the content and format of the test remain true to the original version, even though the scores from the translated tests will not be combined with the scores from the original version, nor will the performance of those taking the different versions be compared. Despite the different intent for the use of scores, these guidelines represent good practice and can be helpful for schools in establishing test translation procedures.

Gross (1986) prepared a manual enumerating the ideal procedures for translating the lactation consultant licensing exam. Gross and Scott (1989) provide an overview of these guidelines in an article in *Evaluation and the Health Professions,* in which they analyze the translation of the exam administered by the International Board of Lactation Consultant Examiners (IBLCE):

> Because of IBLCE's international scope and the probability of testing in languages other than English, the English version should avoid jargon and vernacular and idiomatic phraseology. . . . Traditional item writing guidelines were strictly followed. . . . Translators were directed

to maintain the format of the item stem in the translated version. For example, if the English stem was in the form of an incomplete sentence, the translated stem had to be in the form of an incomplete sentence rather than forming a question. Other issues such as grammatical relationships and verb tense and selection were emphasized also in order to avoid subtle changes in meaning (e.g., "will" versus "would" versus "should"). Finally, translators were asked to avoid making the translated item "more interesting." As an example, the use of synonymous terms (e.g., baby, infant, neonate) interchangeably within the same item was to be avoided because of subtle changes in meaning.

Upon completion of the translation, standard operating procedures required that a different bilingual subject matter expert translate the translated version back to English. This individual received the same guidelines as the initial translator. The retranslated version of the test was then forwarded to a third subject matter expert who was not necessarily bilingual. The responsibility of this third expert was to compare the translated English version with the original English version for corroboration. Any item for which a substantive discrepancy was noted would be flagged for subsequent linguistic review. (p. 66)

Back translation is a common practice in test translation. The back translation process involves checking every word against the original and requires three different translators. It is particularly important that these translators be native language speakers of the language in which the test is translated in order to pick up the cultural nuances as well as nuances in syntax and semantics. Back translation is the accepted procedure of the American Translators Association (ATA). The ATA *Code of Professional Conduct and Business Practices* (1997) recommends that translators have up-to-date knowledge of the subject material and its terminology and mastery of the target language equivalent to that of an educated native speaker.

Auchter and Stansfield (1997) report on a project to translate five forms of the General Educational Development (GED) test into Spanish. The GED is a widely used test designed to enable people who did not graduate from high school to earn the equivalent of a high school diploma. Most colleges and universities, the military, and many employers recognize the validity of the GED. Auchter and Stansfield cite several guidelines that in their view represent best practices in test translation:

- Select those tests and versions of a test most amenable to test translation. The criteria include recency of the test specifications, relevancy of content to Hispanic examinees,

and ease with which the language used in the test could be translated into Spanish.

- Select certified trained translators who are native language speakers.
- Educate translators to use all variants of words or phrases; to be sensitive to issues of dialect and syntax; and to conduct initial forward translation, including compiling a list of items that are difficult to translate or words that reflect cultural bias.
- Examine the translated version against the original to judge the congruity of the translation with the English-language version.
- Conduct additional review of the tests using a contractor who has specific expertise in test construction.
- Conduct yet another review using two additional reviewers selected because of their special expertise in understanding variations in dialect that might influence how the test questions are interpreted by various Spanish language speakers.
- Conduct key verification in Spanish to identify the correct answers.
- Document the process that was used to translate each test.

Auchter and Stansfield also describe issues that arise in translating subject matter content. In their study, subject matter experts were called in to review each of the subject-specific tests. Not surprisingly, the mathematics test provided the fewest translation issues. These authors do not support the use of back translation as does Gross (1986) but rather, as evidenced by their guidelines, they recommend multiple variations of front review.

Another translation procedure is *side-by-side translation.* In this model, the translated version of the test is provided with the English-language version (Anderson, Liu, Swierzbin, Thurlow, & Bielinski, 2002). Anderson and colleagues describe a pilot study in which LEP students received versions of the Minnesota Basic Standards Reading Test in both English and Spanish, on audiocassette and in writing. The scores of students receiving the test in both languages were compared to scores of students who received only the English version. Most of the students reported that they did not use the taped version of the test at all and used the Spanish version to translate specific words from English. The scores of students assigned different versions of the test

were not significantly different; however, the pilot study involved a small sample size, and the methodology seems promising enough to warrant further study.

Some states and large school districts employ the use of professional translation services. These organizations specialize in translating documents, including tests, into different languages. For example, the Center for Applied Linguistics website (www.cal.org/services) indicates that the company can provide translation into and from all the major world languages.

Using Interpreters

Many school districts, individual teachers, and other school professionals rely on informal means to accomplish translations. A common practice is to ask someone already affiliated with the school, such as a parent, a family friend, another professional, or a school staff member to provide translation services (Dale, 1986). Although standardized tests are no longer valid after translation into another language, other kinds of assessments are more amenable to these kinds of informal translation procedures. These assessments may include case histories, oral interviews, and informal teacher-made assessments. Informal translation procedures are also appropriate for interpreting the directions on nonverbal performance tests.

Wyatt (1998) writes that using a family member or family friend as an interpreter has several advantages. The student may be more comfortable with someone familiar, and the interpreter is more likely to speak the same dialect as the student. Wyatt reports disadvantages as well, however, such as the friend or family member trying to help too much. He or she may misrepresent the student's answers in order to present the student in the best possible light or may inappropriately coach the student to perform better.

In addition, regardless of the relationship between the interpreter and the student, untrained interpreters may be prone to mistranslate, not keep up with the student's rate of speaking, forget to include words, or editorialize or elaborate on the student's actual responses. It is critical, therefore, that the interpreter be educated regarding the teacher's expectations and the proper way to administer instructions and collect information (Wyatt, 1998). McCann, Napoli, and Wyatt (1996) found that 40 percent of California school speech-language pathologists who use interpreters are concerned about adequate interpreter training. Wyatt (1998) reports studies that suggest that optimally the interpreter and

the test administrator should meet three times: once to review the client's background and the assessments that will be conducted; next to conduct the actual assessment; and a third time to discuss the interpreter's perceptions of what occurred during the assessment. Test administrators using interpreters can contribute to the accuracy of the translation by speaking slowly and clearly, and by avoiding jargon.

Standard 9.5 of the *Standards for Educational and Psychological Testing* (AERA et al., 1999) addresses the use of interpreters in testing situations:

> When an interpreter is used in testing, the interpreter should be fluent in both the language of the test and the examinee's native language, should have expertise in translating, and should have basic understanding of the assessment process. Although individuals with limited proficiency in the language of the test should ideally be tested by professionally trained bilingual examiners, the use of an interpreter may be necessary in some situations. If an interpreter is required, the professional examiner is responsible for ensuring that the interpreter has the appropriate qualifications, experience, and preparation to assist appropriately in the administration of the test. It is necessary for the interpreter to understand the importance of following standardized procedures, how testing is conducted typically, the importance of accurately conveying to the examiner an examinee's actual responses, and the role and responsibilities of the interpreter in testing. (p. 98)

Problems in Test Translation

Regardless of the quality of a translation, whether performed formally by a professional translation service or informally using an interpreter, there are several other potential problems that can influence the usefulness of translations.

One major variable that influences the utility of translated tests are student characteristics, including attitude. In a National Center on Educational Outcomes (NCEO) study on the impact of bilingual accommodations for LEP students on statewide reading tests, Anderson and colleagues (2002) reported the following findings:

- Accommodations and modifications are not a guaranteed formula for helping LEP students pass a standardized test.
- Translations are not appropriate for every speaker of a particular language.
- Not every student wants, or will use, an accommodation involving translation on a high-stakes test.
- A standardized means of determining which students are likely to benefit from translations should be created.
- English language proficiency, native language proficiency, level of test anxiety, and level of peer pressure to use an English version of a test contribute to determining which students may benefit from a translated test.
- Even within a single language group, the ability to benefit from test translation varies from student to student; generalizations based only on linguistic background should not be made.

The number of different languages spoken by schoolchildren in many states and school districts makes the concept of complete translations fiscally and pragmatically unfeasible even if the psychometric challenges regarding test validity, reliability, cultural bias, and population norming can be overcome.

Additionally, there are insufficient numbers of teachers and other school personnel who are trained to administer, score, and interpret tests in other languages. This is a particular problem for individualized assessment procedures such as those performed by school psychologists, speech-language pathologists, or learning disability specialists because these procedures require a great deal of interaction between the test administrator and the student.

Alternatives to Test Translation

Test translation problems have, in fact, made test translation an unpopular accommodation. In a survey of accommodations employed by states for linguistic minority students, test translation ranked low (Liu, Thurlow, Spicuzza, & Heinze, 1997). Several other methods of accommodating students who are bilingual or who have limited English proficiency exist: performance rating scales, nonverbal measures (particularly of intelligence), tape-recorded test instructions in the student's native language, and allowing additional time to complete the assessment. These accommodations also have advantages and disadvantages. In fact, they are subject to the same issues regarding

 BEST COPY AVAILABLE

reliability, validity, and norming as are accommodations using test translation. Standard 11.9 of the *Standards for Educational and Psychological Tests* (AERA et al., 1999) addresses the issue of using accommodations that do not compromise the reliability, validity, or norms of a test:

> When a test user contemplates an approved change in test format, mode of administration, instructions or language used in administrating the test, the user should have a strong rationale for concluding that the validity, reliability and appropriateness of norms will not be compromised. (p. 115)

Alternatives to test translation are often difficult to implement because they require subjective interpretations by the examiner. As a result they are more time-consuming to score and harder to norm. Additionally, teachers are most familiar and comfortable with paper-and-pencil tests because these tests are the mode of assessment that teachers themselves probably experienced in their own educational careers.

Regardless of what accommodations are available, the issue of whether to provide accommodations at all and, specifically, when is it appropriate to translate a test versus using some the accommodations noted previously is often a difficult decision. Standard 9.10 of the *Standards for Educational and Psychological Testing* (AERA et al., 1999) provides guidelines for testing language proficiency:

> Inferences about test takers' general language proficiency should be based on tests that measure a range of language features, and not on a single linguistic skill.
> For example, a multiple-choice, pencil-and-paper test of vocabulary does not indicate how well a person understands the language when spoken or how well the person speaks the language. (p. 99)

Furthermore, Standard 9.3 states:

> When testing an examinee proficient in two or more languages for which the test is available, the examinee's relative language proficiencies should be determined. The test generally should be administered in the test taker's most proficient language, unless proficiency in the less

proficient language is part of the assessment.

Unless the purpose of the testing is to determine proficiency in a particular language or the level of language proficiency required for the test is a work requirement, test users need to take into account the linguistic characteristics of examinees who are bilingual or use multiple languages. This may require the sole use of one language or use of multiple languages in order to minimize the introduction of construct-irrelevant components to the measurement process. For example, in educational settings, testing in both the language used in school and the native language of the examinee may be necessary in order to determine the optimal kind of instruction required by the examinee. Professional judgment needs to be used to determine the most appropriate procedures for establishing relative language proficiencies. Such procedures may range from self-identification by examinees through formal proficiency testing. (p. 98)

Determining Eligibility for a Translated Test

Large-scale testing programs used by states and school districts generally have specific guidelines to determine which students should be assessed in a language other than English. Sometimes these guidelines are based on practical matters such as cost or feasibility of obtaining alternative language tests.

For example, one major testing company that investigated creating licensing exams in a variety of languages determined that it was not fiscally feasible to do so. The agencies using the licensing exams, the test takers, and the testing company were none of them in a position to support the translation costs. Additionally, there were concerns that offering translated versions in some languages but not others might appear to be discriminatory. Instead, the testing company created a policy that permits individuals whose primary language is not English to apply for an accommodation that allows them time and a half to complete the exam. Preliminary research by the testing company has shown that this accommodation doesn't statistically improve performance on the exam. Other research, reported by Anderson and colleagues (2002) and by Ascher (1990), demonstrates that limited English speakers often need more time to take tests because of the additional time they require for language processing—that is, internally interpreting test items from one language to another.

Policies on who may use a translated test are often based on the number of years of English instruction a person has received, rather

than on an individualized assessment of English language proficiency. An example of this kind of policy is illustrated by the language in the recent ESEA reauthorization mentioned earlier that requires inclusion of all students in large-scale testing programs if they have been instructed in English for more than three years.

Indeed there appears to be no normative definition of what constitutes limited English language proficiency, much less what proficiency level, or lack of proficiency, provides the student with the right to receive appropriate test translations and related accommodations. Several terms, including linguistic or language minority, limited English proficient, and bilingual, are generally used to describe students whose primary language is not English, but no universal normative definitions are used by all the states. (Liu et al., 1997). Liu and colleagues report in their review of the literature on LEP students and assessment that each state has a definition but that the state definitions contain different components, generally variants of the federal definition.

The federal definition from Title VII of the Improving Schools Act of 1994 (PL 103-382, Part E, Section 7501: Definitions, Regulations) defines a student as LEP if he or she meets the following criteria:

> A student that has sufficient difficulty speaking, reading or understanding the English language and whose difficulties may deny such individual the opportunity to learn successfully in classrooms where the language of instruction is English or to participate in society due to one or more of the following reasons:
>
> - Was not born in the United States or whose native language is a language other than English and comes from an environment where a language other than English is dominant;
> - Is a native American or Alaskan native . . . and comes from an environment where a language other than English has had a significant impact on such individual's level of English proficiency; or
> - Is migratory and whose native language is other than English and comes from an environment where a language other than English is dominant.

Special Considerations in Test Translation

Other considerations in test translation involve distinguishing a language issue from an education issue or learning disability and understanding that one translation does not suit every speaker of a particular language.

Language or education deficiency? Some students with LEP may have come from backgrounds where formal education was limited. This may be particularly true for political refugees from nations whose schools were closed due to military or civil unrest. It may also be true of immigrants arriving from nations where certain groups are denied access to education because of their ethnicity or their sex. For individuals with limited schooling, translating tests into their native language will not help them perform comparably to other students. Translation can provide a better gauge of their educational level, however, and a sense of how much of the difficulties they may be experiencing are due to language differences versus educational differences.

Language difference or disability? Some students with LEP also have a language or learning disability. It is important, therefore, to test the student in his or her native language to determine any special education needs (ASHA, 1985). Federal special education laws (e.g., IDEA) mandate testing in the student's native language; however, disproportionately large numbers of linguistic minority students are mistakenly labeled as having a disability and are assigned to special education programs. Special care must be taken to assess students appropriately so that they are neither inappropriately denied nor inappropriately placed in special education programs. Again, the proper use of test translations, especially the use of skilled interpreters who also have knowledge of appropriate linguistic and behavioral norms, can be invaluable in ensuring that students are properly diagnosed and educated.

The fallacy of the "one translation fits all" model. Much has been written about variations in test performance among English-language speakers due to linguistic differences. Differences in geographic, social, ethnic, and racial background as well as other demographic variables contribute to differences in language use. Differences in language use can include differences in vocabulary, syntax (word order in a sentence), morphology (use of word endings), grammar, pronunciation, and

cultural referents. Large-scale testing programs commonly use techniques such as differential item functioning analysis (DIF) to control for any bias in the content of a question that may be a result of these linguistic and cultural differences. Similar variations in language use among speakers of the same language occur in most languages. Test translations must be sensitive to dialectal and other variations that may occur among common language speakers. One approach is to use words that are expected to be understood by all; another is to identify and incorporate several variants of words in the translations. For oral translations, identifying interpreters from the same geographic region and social and cultural background as the student is very important in contributing to accurate and appropriate translations.

Summary

Test translations can be useful for many students who are LEP and determined to be eligible for testing accommodations. Several factors, however, influence the utility of test translations.

One set of factors relates to individual student characteristics and needs. These include the students' native language proficiency, dialect, and culture, as well as the student's interest in using a translated test. Not all students will benefit from test translation. Nor will all translations in a particular language be appropriate for every student who speaks that language.

Another set of factors relates to the technical features of the test, including item bias, validity, and norming. Test users must be sure not only that the test items represent an accurate linguistic translation but also that the cultural referents are appropriate for the individual student's background. Appropriate methods of translation, including back translations or multiple forward translations, must be used. Additionally, the test must be validated and normed on a linguistically and culturally appropriate population, that is, a population similar in demographics to that of the student's. Lastly, teachers, translators, and others involved in the testing process must be educated on how to select, administer, score, and interpret translated tests.

References

AERA, APA, & NCME. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

AHSA [American Speech-Language-Hearing Association]. (1985, June). Clinical management of communicatively handicapped minority language populations. *Asha, 27*(6), 29–32.

Anderson, M., Liu, K., Swierzbin, B., Thurlow, M., & Bielinski, J. (2002, August). *Bilingual accommodations for limited English proficient students on statewide reading tests, phase 2* (National Center for Educational Outcomes, Minnesota Report 31). Retrieved March 11, 2002, from http://education.umn.edu/nceo/OnlinePubs/MnReport31.html.

Ascher, C. (1990). *Assessing Bilingual Children for Placement and Instruction* (ERIC/CUE Digest No. 65). New York, NY: ERIC Clearinghouse on Urban Education. (ERIC Document Reproduction Service No. ED322273)

ATA [American Translators Association]. (1997, November). *Code of professional conduct and business practices.* Retrieved February 21, 2003, from www.atanet.org/codeofprof.htm.

Auchter, J. & Stansfield, C. (1997). *Developing parallel tests across languages: Focus on the translation and adaptation process.* Paper presented at the annual Large Scale Assessment Conference, Colorado Springs, CO. (ERIC Document Reproduction Center No. ED414320)

Battle, D. E. (1998). *Communication disorders in multicultural populations* (2nd ed.). Boston, MA: Butterworth-Heinemann.

Bracken, B. A., & McCallum R. S. (1999). International testing: The universal nonverbal intelligence test. *The International Test Commission Newsletter 9*(1), 7–11.

Council of Chief State School Officers. (2001). *Annual survey of state student assessment programs for 1998–99 school year*. Washington, DC: Author.

Dale T. C. (1986). *Limited-English-proficient students in the schools: Helping the newcomer.* Washington, DC: ERIC Clearinghouse on Languages and Linguistics. (ERIC Document Reproduction Center No. ED279206)

Figueroa, R. A. (1990). Best practices in the assessment of bilingual children. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (vol. 11). Washington, DC: National Association of School Psychologists.

Geisinger, K. F., & Carlson, J. F. (1992). Assessing language-minority students. *Practical Assessment, Research & Evaluation, 3*(2).

Gross, L. J. (1986). *Examination translation guidelines* (monograph). Memphis, TN: International Board of Lactation Consultant Examiners.

Gross, L. J., & Scott, J. W. (1989). Translating a health professional certification test to another language: A pilot analysis. *Evaluation and the Health Professions, 12*(1).

International Test Commission. (1999). *International guidelines for test use: Version 2000.* Stockholm: Author.

Liu, K., Thurlow M., Spicuzza, R., & Heinze, K. (1997). *A review of the literature on students with limited English proficiency and assessment* (National Center on Educational Outcomes, Minnesota Report 11). Retrieved February 21, 2003, from http://education.umn.edu/nceo/OnlinePubs/MnReport11.html.

McCann, M., Napoli, M., &Wyatt, T. (1996). *Use of paraprofessionals with low-incidence language populations: A survey.* Presentation at the California Speech and Hearing Association conference, Monterrey, CA.

National Academy of Sciences. (2002). *Report on minority students in special and gifted education.* Washington, DC: Author.

Stansfield, C. W. (1996). Content assessment in the native language. *Practical Assessment, Research & Evaluation, 5*(9). Available online at http://ericae.net/pare/getvn.asp?v=5&n=9.

Wyatt, T. (1998). Assessment issues with multicultural populations. In D. Battle (Ed.),*Communication disorders in multicultural populations* (2nd ed.). Boston, MA: Butterworth-Heinemann.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[ ] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[X] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").